

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA 147000022596013 /altid=gi 10438148 /def=dbj BAB15179.1 (AK...	931	0.0
CRA 1000682322899 /altid=gi 5802604 /def=gb AAD51732.1 (AF1752...	795	0.0
CRA 335001098690982 /altid=gi 11436851 /def=ref XP_003547.1 UD...	679	0.0
CRA 18000005226060 /altid=gi 5803213 /def=ref NP_006789.1 UDP ...	677	0.0
CRA 18000005155438 /altid=gi 4507821 /def=ref NP_001068.1 UDP ...	676	0.0
CRA 18000005147363 /altid=gi 6175083 /def=sp P06133 UDB4_HUMAN ...	675	0.0
CRA 18000004953169 /altid=gi 484383 /def=pir JN0619 glucuronos...	674	0.0
CRA 18000005148770 /altid=gi 3153832 /def=gb AAC95002.1 (AF064...	674	0.0
CRA 18000004903910 /altid=gi 4507819 /def=ref NP_001067.1 UDP ...	669	0.0
CRA 18000005164461 /altid=gi 3426332 /def=gb AAC32272.1 (AF081...	668	0.0
CRA 1000682327923 /altid=gi 5881246 /def=gb AAD55093.1 AF180322...	668	0.0
CRA 18000005219476 /altid=gi 8134780 /def=sp Q9XT55 UDBJ_MACFA ...	667	0.0

BLAST dbEST hits:

gi 10971169 /dataset=dbest /taxon=96...	383	e-103
gi 11068678 /dataset=dbest /taxon=96...	234	6e-59
gi 679005 /dataset=dbest /taxon=9606 /...	212	2e-52
gi 3173232 /dataset=dbest /taxon=9606 ...	212	2e-52
gi 3134358 /dataset=dbest /taxon=9606 ...	212	2e-52
gi 10298020 /dataset=dbest /taxon=96...	200	8e-49
gi 11974507 /dataset=dbest /taxon=96...	196	1e-47
gi 11973717 /dataset=dbest /taxon=96...	172	2e-40
gi 12673874 /dataset=dbest /taxon=96...	137	1e-29
gi 10887798 /dataset=dbest /taxon=96...	125	4e-26

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|10971169 Kidney-hypernephroma
gi|11068678 HepG2 cell line
gi|679005 Liver
gi|3173232 Kidney
gi|3134358 Kidney
gi|10298020 Hepatocellular carcinoma
gi|11974507 Normal pigmental retinal epithelium
gi|11973717 Normal pigmental retinal epithelium
gi|12673874 Kidney hypernephroma
gi|10887798 Kidney

Expression information from PCR-based tissue screening panels:

Human fetal liver

1 MRSDKSALVF LLLQLFCVGC GFCGKVLVWP CDMSHWLNK VILEELIVRG
 51 HEVTVLTHSK PSOLIDYRKPS ALKFEVVHMP QDRTEENEIF VDLALNVLP
 101 LSTWQSVIKL NDFVFEIRGT LKMMCESFIY NQTLMKKLQE TNYDVMLIDP
 151 VIPCGDLMAE LLAVPFVLT L RISVGGNMER SCGKLPAPLS YVPVPMGTGLT
 201 DRMTFLERVK NSMLSVLFHF WIQDYDYHFW EEFYSKALGR PTTLCTVGVK
 251 AEIWLIRTYW DFEFPQPYQP NFEFVGGLHC KPAKALPKEM ENFVQSSGED
 301 GIVVFSLGSL FQNVTEEKAN IIASALAQIP QKVLWRYK GK KPSTLGANTR
 351 LYDWIPQNDL LGHPKTKAFI THGGMNGIYE AIYHGVPMVG VPIFGDQLDN
 401 IAHMKAKGAA VEINFKTMTS EDLLRALRTV ITDSSYKENA MRLSRIHHDQ
 451 PVKPLDRAVF WIEFVMRHKG AKHLRSAHD LTWFQHYSID VIGFLLTCVA
 501 TAIFLFTKCF LFSCQKFNKT RKIEKRE (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
 N-glycosylation site

Number of matches: 3

1 131-134 NQTL
 2 313-316 NVTE
 3 518-521 NKTR

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
 cAMP- and cGMP-dependent protein kinase phosphorylation site

Number of matches: 2

1 67-70 RKPS
 2 340-343 KKPS

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site

Number of matches: 6

1 3-5 SDK
 2 120-122 TLK
 3 169-171 TLR
 4 200-202 TDR
 5 435-437 SYK
 6 520-522 TRK

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site

Number of matches: 9

1 62-65 SLID
 2 141-144 TNYD
 3 204-207 TFLE
 4 243-246 TLCE
 5 258-261 TYWD
 6 296-299 SSGE
 7 297-300 SGED
 8 419-422 TSED
 9 435-438 SYKE

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
 Tyrosine kinase phosphorylation site

Number of matches: 2

1 122-130 KMMCESFIY
 2 136-143 KKLQETNY

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 4

1	19-24	GCGFCG
2	276-281	GGLHCK
3	373-378	GGMNGI
4	377-382	GIYEA

[7] PDOC00009 PS00009 AMIDATION
Amidation site

338-341 KGKK

[8] PDOC00359 PS00375 UDPGT
UDP-glycosyltransferases signature

354-397 WIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQ

[9] PDOC00804 PS01047 HMA
Heavy-metal-associated domain

12-41 LLQLFCVCGFCGKVLVWPCDMSHWLNVKV

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	5	25	1.802	Certain
2	157	177	0.765	Putative
3	181	201	0.779	Putative
4	377	397	0.735	Putative
5	491	511	1.931	Certain

BLAST Alignment to Top Hit:

>CRA|147000022596013 /altid=gi|10438148 /def=dbj|BAB15179.1|
(AK025587) unnamed protein product [Homo sapiens]
/org=Homo sapiens /taxon=9606 /dataset=nraa /length=449
Length = 449

Score = 931 bits (2381), Expect = 0.0
Identities = 448/449 (99%), Positives = 448/449 (99%)

Query: 79 MPQDRTEENEIFVDLALNVLPLGLSTWQSVIKLNDFVEIRGTLKMMCESFIYNQTLMKKL 138
MPQDRTEENEIFVDLALNVLPLGLSTWQSVIKLNDFVEIRGTLKMMCESFIYNQTLMKKL
Sbjct: 1 MPQDRTEENEIFVDLALNVLPLGLSTWQSVIKLNDFVEIRGTLKMMCESFIYNQTLMKKL 60

Query: 139 QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG 198
QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG
Sbjct: 61 QETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGNMERS CGKLPAPLSYVPVPM TG 120

Query: 199 LTDRMTFLERVKN SMLS VLFHFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT 258
LTDRMTFLERVKN SMLS VLFHFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT
Sbjct: 121 LTDRMTFLERVKN SMLS VLFHFWIQDYDYHFWEEFYSKALGRPTTLCETVGKAEIWLIRT 180

Query: 259 YWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLSLFQNVTEEK 318
YWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLSLFQNVTEEK
Sbjct: 181 YWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSSGEDGIVVFSLSLFQNVTEEK 240

Query: 319 ANIIASALA QIPQKVLWRYKGKKPSTLGANTRYDWIPQNDLLGHPKTKAFITHGGMNGI 378
ANIIASALA QIPQKVLWRYKGKKPSTLGANTRYDWIPQNDLLGHPKTKAFITHGGMNGI
Sbjct: 241 ANIIASALA QIPQKVLWRYKGKKPSTLGANTRYDWIPQNDLLGHPKTKAFITHGGMNGI 300

Query: 379 YEAIYHGVP MVGVPIFGDQLDNIAHMAKAGAAVEINFKTMTSEDLLRALRTVITDSSYKE 438
YEAIYHGVP MVGVPIFGDQLDNIAHMAKAGAAVEINFKTMTSEDLLRALRTVITDSSYKE
Sbjct: 301 YEAIYHGVP MVGVPIFGDQLDNIAHMAKAGAAVEINFKTMTSEDLLRALRTVITDSSYKE 360

Query: 439 NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLLTC 498
NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLL C
Sbjct: 361 NAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTW FQHYSIDVIGFLLAC 420

Query: 499 VATAIFLFTKCF LFSCQKFNKTRKIEKRE 527 (residues 79-527 of SEQ ID NO:2)
VATAIFLFTKCF LFSCQKFNKTRKIEKRE
Sbjct: 421 VATAIFLFTKCF LFSCQKFNKTRKIEKRE 449 (SEQ ID NO:4)

>CRA|1000682322899 /altid=gi|5802604 /def=gb|AAD51732.1| (AF175221)
UDP glucuronosyltransferase UGT2A3 [Cavia porcellus]
/org=Cavia porcellus /taxon=10141 /dataset=nraa
/length=530
Length = 530

Score = 795 bits (2030), Expect = 0.0
Identities = 377/530 (71%), Positives = 435/530 (81%), Gaps = 3/530 (0%)

Query: 1 MRSDKSALV FLLLQLFCVGCFCGKVLVWPCDMSHWLNVKVILEELIVRGHEVTVLTHSK 60
M K A LLL L C G GFCGKVLVWPC+MSHWLN+K +LEEL+ RGHEVTVL T S
Sbjct: 1 MAPGKLASAV LLLLCCAGSGFCGKVLVWPCMSHWLNLKTLEELVVRGHEVTVL TSN 60

Query: 61 PSLIDYRKPSALKFEVVHMPQDRTEENEI---FVDLALNVLPLGLSTWQSVIKLNDFVEI 117
IDY + A FEV+ +P D+ I F++LA+NV+P+ WQS L FFV+I
Sbjct: 61 NLFIDYNRHPAFNFEVIPVPTDKNMSSENILNEFIELAVNVMPMTPLWQSGKLLQQFFVQI 120

Query: 118 RGTLMKMMCESFIYNQTLMKKLQETNYDVMLIDPVI PCGDLMAELLAVPFVLT LRISVGGN 177
L + C + +YNQ+LMKKL+++ YDV++ DPVIPCGL+L+AE+L VPFV L+ S+G
Sbjct: 121 TEDLGLNCRNTRVYNQSLMKKLRLRDSKYDVLVTD PVI PCGELVAEMLGVFVNMLKFSMGHT 180

Query: 178 MERSCGKLPAPLSYVPVPMGTGLTDRMTFLERVKNSMLSVLFHFWIQDYDYHFWEEFYFYSKA 237
 +E+ CG+LPAP SYVPVP+ GLT RMTF+ERVKN + SVLF FWIQ YDY FW++FYS+A
 Sbjct: 181 IEKYCGQLPAPPSYVPVPLGGLTTRMTFMERVKNMVFSVLFDFWIIQQYDYKFDQFYSEA 240

Query: 238 LGRPTTLCETVGKAEIWLIRTYWDFEFPQPYQPNFEFVGGLHCKPAKALPKEMENFVQSS 297
 LGRPTTLC +GKAEIWLIRTYWDFEFP+PY PNFEFVGGLHCKPAK LPKEME FVQSS
 Sbjct: 241 LGRPTTLCIMGKAEIWLIRTYWDFEFPYPYLPNFEFVGGLHCKPAKPLPKEMEEFVQSS 300

Query: 298 GEDGIVVFSLSLGFQNVTEEKANIIASALAQIPQKVLWRYKGKKPSTLGANTRYLDWIPQ 357
 GEDG+VVFSLGS+ +N+TEEKAN+IASALAQIPQKVLWRYKGKKP+TLG NTRL+DWIPQ
 Sbjct: 301 GEDGVVVFSLGSMVKNLTEEKANLIASALAQIPQKVLWRYKGKKPATLGPNTRLFDWIPQ 360

Query: 358 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPVMGVPIFGDQLDNIAHMKAKGAAVEINFKT 417
 NDLLGHPKTKAFITHGG NGIYEAIYHGVPVMVG+PIF DQ DN+A MKAKGAAVE+N T
 Sbjct: 361 NDLLGHPKTKAFITHGGSNGIYEAIYHGVPVMGMPIFSDQPDNLAMKAKGAAVEVNMNT 420

Query: 418 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 477
 MTS DLL ALRTVI D +YKENAM+LSRIHHDQPVKPLDRA FW+EFVM HKGAKHLR A
 Sbjct: 421 MTSADLLGALRTVINDPTYKENAMKLSRIHHDQPVKPLDRAAFWVEFVMHHKGAKHLRVA 480

Query: 478 AHDLTWFQHSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 527 (residues 1-
 527 of SEQ ID NO:2)
 AHDL+WFQ++S+DVIGFLL CVA+AI L TKC LFS Q F K K K+E
 Sbjct: 481 AHDLSWFQYHSLDVIGFLLACVASAILLVTKCCLFSFQNFQIKIGKRIKKE 530 (SEQ ID NO:5)

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00201	UDP-glucuronosyl and UDP-glucosyl transferas	962.0	1.5e-285	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00201	1/1	24	525 ..	1	507 []	962.0	1.5e-285

```

1 TTCTAGAGGG TTGGAACAAC TTTTCCCTGA TACATTGCAT TTTTGTGATA
51 CCTTCAGTAC ATGTTAAACT GGCAACCACC AGTGAACCTT ACTCTTAAAA
101 TATTAATTTT TAACCTCTGT GCTTATATTG TCATTCAAC TCCTTGCTTA
151 GTAACCTACAA AACCATTTGCA GATCAGTGTG TGAGGGAAC GCCATCATGA
201 GGTCTGACAA GTCAGCTTTG GTATTTCTGC TCCTGCAGCT CTTCTGTGTT
251 GGCTGTGGAT TCTGTGGGAA AGTCCTGGTG TGGCCCTGTG ACATGAGCCA
301 TTGGCTTAAT GTCAAGGTCA TTCTAGAAGA GCTCATAGTG AGAGGCCATG
351 AGGTAACAGT ATTGACTCAC TCAAAGCCTT CGTTAATTGA CTACAGGAAG
401 CCTTCTGCAT TGAAATTTGA GGTGGTCCAT ATGCCACAGG ACAGAACAGA
451 AGAAATGAA ATATTTGTTG ACCTAGCTCT GAATGTCTTG CCAGGCTTAT
501 CAACCTGGCA ATCAGTTATA AAATTAAATG ATTTTTTTGT TGAAATAAGA
551 GGAACCTTAA AAATGATGTG TGAGAGCTTT ATCTACAATC AGACACTTAT
601 GAAGAAGCTA CAGGAAACCA ACTACGATGT AACGCTTATA GACCCTGTGA
651 TTCCCCGTGG AGACCTGATG GCTGAGTTGC TTCCAGTCCC TTTTGTGCTC
701 ACACCTAGAA CTTCTCTAAG AGGCAATATG GAGCGAAGCT GTGGGAAACT
751 TCCAGCTCCA CTTTCTATG TACCTGTGCC TATGACAGGA CTAACAGACA
801 GAATGACCTT TCTGGAAAGA GTAAAAAATT CAATGCTTTC AGTTTGTTC
851 CACTTCTGGA TTCAGGATTA CGACTATCAT TTTTGGGAAG AGTTTATAG
901 TAAGGCATTA GGTAAGACAC TTTTGTTTTA TTTTAAATTT AGTTATCAAA
951 AGAAATATTT TTAATAATTG TCATACATTG TCTATGACAT ATATATGCAG
1001 GTCAATGAGT TTTTGTAGAA AATGTTGTAG CTGTTTTTCA TAAAGAAAGT
1051 GTATTTGTTC TAAGCGTAAG ATAACCTACT TTCTTAATAC CAGTAATATA
1101 CTTAAAAATG ATCATCAATA ACTAAGAGAT TATATTTTGT ATTTCTCTCA
1151 AATAGCGCAA ATCAACATCA CATATTTTGT AGAATCACTG ATTGTAGTC
1201 TGAATGTTAT AGAATTTCTA TTGAAATAAA ATGCTAATCA TTATTTCTC
1251 TCTCATCATG TATTTAAGAA AATCTTCAGA AGGTCTTCTT TGAATTAATT
1301 TTTCAAGAGT CATTAATTTG AACATTTTCT AGAATCTTTT AATTTCTTAG
1351 GTGATTACTT CACAAAAACT TGAAAAAATA TTATAAAAAG TTAATAAACT
1401 TACGGTCTTG TGGGGCATAA GATAGTAGAA TTTTACTTT ACTGATATAC
1451 ACCTATTTGA CTTATTTTTT TTTCTTTGCT TTAGTGATAA AAAGTTGTTT
1501 TGCTTTGCAA TTTTCATATA GTTGTGATCA GAGCTGGTCA ATGCAAGACA
1551 TGTTTTTATC CAAATATGTT TGAGAATTAT GTAGAAACAT GAAAAAAGGT
1601 ACAATTATAT CCGACACTAA AATATTGTTT AATGTATTCC AACGAATTCT
1651 TATGCATAGA CTGTTTCACA GAACATAAT TCAGAGGATC CCAGTTCAAA
1701 TGTCCTTAGC CTTAGACATG ATTTGAATTT ACATATATTG ATTTGCTTTA
1751 AATAATTTTC CATTAGTAA GCTGTGCCTA GCTGCAGATA GCCTACCAGG
1801 CTTTATGGAT CTAGGTAAC AATACAAATC TCTTGGCCTC AAGTCTACAT
1851 TCAGATATTA ATTTAAAGGG GTACAGCTAT ATAGAGGTCA CTGGCAAATT
1901 TTGGTAAAAT AGGATTATAG TAAAAGCCCC CTGACAAGAT TGAAATTTAA
1951 AATAAAACAA AAGTGTATC AAAGGGGTGA AAGAGCATTT TCCAATAAAC
2001 AAAAGTGGGT TCTGGCCATG CATTAGAAA TTCCCCAACA ATTCTTTAAA
2051 AATCATGGAG CAGCTTGATA TATAAGAAAT TCATTTAATA ACTATATTTA
2101 TTATGTAGCT CCAACTTACT AAATTATTGA TTATTATATA TTTTATAGAA
2151 TTATCTATTG TTGATCTAAA TCAAGAGTAT ATATTCAAAC AACTATAGGA
2201 AAAGGGATAT CAGTCAATTT CAATTCAAGG ATTTATTTCC ATAAGTGCTT
2251 ACGCACAGGT GTATTTTATT TTATTATACA TTGCTTTATT GTCCTTCACA
2301 AAAATTGCAA TTTACAAATT AAAGGTTTTT GAAAACCTTG AATCAAGCTA
2351 ATCAATTTGG CGTAATATTT CCAACAACAA GTGTGTACTT TTGACTCTAT
2401 CACATATTGG CATTATCAT GCTTTTCAA ATTTTTCATT GTTATATCTG
2451 TTACGGTGAT CTGGGATCAG TGTTCTTTGA TGGTTACACG TTTATTAGCT
2501 TGGGGGCACC TTGATGTGTT ACAATATAAG ACAGCAAAC TAATTATAAA
2551 TGTTGTGCAT GTACTAACTG CTCCGCTGAT TCGTTTCCCC ATCCCACTTC
2601 TTCTTAGGCC TCCCTATTCC CTGAGACACA GTAATATAAC ATACAATGAC
2651 TTCTAAATGT TCCAGTGAAA AGAAAAGTAG CAGGTCTCTC AATTTAAACC
2701 AAAAATATAA AGGAATAAGT TTAATGAGTA CTATAGTTTA GATATGGTTT
2751 GCTTGACCTT ACAAAATCCT GTGTTGAAAT TTGATCACC AATTGGAGG
2801 TGGGGCTTGA TGGGAAGTGT TAGGGTCATG AGGGTAGATT CCTATGAGT
2851 ACATTAATGC TCTCCCTGGG GAAATGGGTG AGTTCGTTCT CACTCTATTA
2901 GGTCCCAGGA GAGATAATTA TTAATAAGAG CCAGGAACAT CCACCTTCTT
2951 TCTCTTGCA ATCTCTCATT ATCTGATCCC TGCATTGCT GGCTCCCAAC
3001 ATCTTCTTCA ATGAGTGGAG GAAACCAGAG GTCTTCACCA GACACAGATG
3051 TTGGTGCCAT GCCTCTTGTA TACCCTGAAG AATTGTGAGC CAAATAAAAA
3101 CCTTTTCTT TTACAAATTA GACAGCCTCA GTTATTCCTT TGTAGCAACA

```

FIGURE 3, page 1 of 11

3151 AAAAAAGCCT GGGACAGGCC AAAAATAACA CCATTGCACC AAACAGTTAA
3201 ACAAGATGTG AGTGCAAAAG AAAAGTTTTT GGAGGAAATT AAAAGTGCTA
3251 CTCCAGTGTG CATACAAATG ATAAGAACA ATAACCATTA TCAGTGCTGA
3301 TATGGAGAAA ATTTTAGTTG TCTGGAGAGA AAATCAAATT AGCTAGCCAG
3351 CTGCAGTGAT TCATATCTGT AATCCCAGTA ACTTGGGAGG CTCAGGTGGG
3401 AGAACGGCTT GAGCCCAGAA GTTTGAAGTC CAAGGCTGCA GTGAGCTATG
3451 ATTGCTCCAC TGCACCTCAA CCTAGGTGAT AGAGCAAAAC CACTACCAAA
3501 AAAAAAAAAA AAAAAAGAA GAAAAAGAAA AGAAAAAAA TTAACCAAC
3551 CACAACATCA CCTTAGGTTT TGGCATTAGC TAAAACTAA TACATAGTAA
3601 AGCGTTAACT ATTCATTGTC ATGAAACCTC AGAGAGGAGA GGAAGATGCA
3651 GAAAAAAGA CTGAAGCTAG TAGAGGTTGA CTAATGAGGT TTACAGGAAT
3701 AAATGCCTA CATGATGCAA AAGTTCAATG TGAAGCAATA GGAAGTCATG
3751 CAGAAGACTT AGCTAATATA CTCAGTAAAT GTGGCTACAG TAAACAAATG
3801 ATTTTCAATG TAGACCTAAC AGCCTTCTGT TGGAGAAGA TGCCATTTAA
3851 AACTTTCATA GCTAGAGAAG AGAAGTCAAT GCTTGTCTCT GAAGCTACAA
3901 AAAACAGGCT GAATCTCTTG TAGTGGCTAA TGCAGCTGAT GACAAAGGTA
3951 AAGCCTAATG CCATTACTT TTTGTAATAA TTATAGAGA CTCTTAATAA
4001 TTATGTTAAA TCTACTTTGC CTGTGTTATA TCAATGGAAC AACAAAGCCT
4051 GGATGATATC ACATTGGTAT ATGACATGGC TTATTGAATA TTTTAAGCAC
4101 ACTGTTGAGA CCTATTGCTC AAAAAAGAGG ATTCCTTTCA AAATATTGCT
4151 GCTCATTGAC AATTCACATG GTCAACAAAG GGCTCTGATT AAGATGTACA
4201 GATATTAATG TTTGCCCTGCT TGCTATTATT ACATCCATCT TACATGCCAT
4251 GGATCATATA GCCTTGACTT TCAAGTCTTA TGTAAGAAAT ATATTTTGTA
4301 AGGCTATAGC TCTTACTAAT GGGGAAAGTA TATTGAAAC CTTTCAAAA
4351 GGATTTTTC TTTAGATTG CATTAGAAGC ATTCATGGTT CATGAGAGGA
4401 AGTCAAGATA TTAACATTAA CAAGAGTTTG GAAAAAATTT GATTCTAACT
4451 CTCCTGGATG ATTTTGAGGG ATTGAAGACA TCATGTGAAG AATTAAGTGG
4501 GGATGGGGTG GTCATGAAAA AATAAATAGA ATTATAAGTG GGCTGAAGG
4551 TTTGTCTAAA TTGCTATAAT ATCATGATAA AACTAAAACC TGTAACCCG
4601 GTGAGGAGGT GCTTTTAA CAGTTACTTT TTATAGATGA ACACAGAAAT
4651 TGGTTTTGTG AGTTGGAATC TTCTCCGAGT GAAAAATGCTA TGAACATTGT
4701 TGAAATGGCT ACAAATGACT TAGAATATTA CACAAAATTA GTAGATAAGG
4751 CAGCATCAAG GTTTGAGAGA ATGGACTCAA ATTTTGAAAG AAATCTACT
4801 ATGGGTAAAC TGCTGTGAAA CATCATCATA TGCTACAGAG AAATCTTTCA
4851 TGAAAAGATG AGTCAATTCA TGCAACAATC TTTGTGTCT AATTTTAAAA
4901 ATTGTCCAGC TGCCCTGATC AATCAACAGT AATCAGCACT GAGGCAAGAC
4951 CCTACACAGC AAAAAATAA AATAAAAAAC CTCACTTGCT GAAGACTCAG
5001 CTTATTATTA GCACCTTTTA GCCATACTTT TAACTAAGGT ATGTGCATT
5051 CTTTTTAAAC GTGATGATAT TGCACAGCTA ATAGCCTACA AGGTATGGTT
5101 AACATAACTT TTATATGTCC TGGGACCCAA ATTTGTGTGA ATCACTTTAT
5151 TGACATATTC CTTTTATTGA GATGAAGTGC AACTTATCTT GCAATATCTC
5201 CAAGATATGT GTGTATGGCA TTTCAAATAA GATGTGAAAT TATTTTATTA
5251 GTATAAAAAG CAAATTTAAT TTTCTTTCCT TTGATCATCT TTATCCTTGT
5301 TACTGTGTAT TATCCTTTA AACATTGAAT GACTCCAATT GTTTAAACT
5351 GAGTCTTCT TAAATGAGTC CTAATATCAT AGTAATTAAA ATCACCTACA
5401 AGTTGGTAAT GCAGGCAGCA TGTGAGGCAC AGAAAAACA AAATTTATAA
5451 GACATAAATG CATTGCTTGA GAAGCTGAGA GAAGGCTCTA TTCTAATTTT
5501 TGATAACTTC AAATGAGTA TCTTCAGTAA AATTTATTCA CTATCAAAT
5551 CAAGGCGTTT GGATTTATGA CCTAGGAAAA AACTTCAAAC ATTAAATGT
5601 GATGACCTTA AAAAGAGGCT CTCCACACTA TGGTGTATAA CACCACAAC
5651 TTTGATTAGA ATTTTAAAGA GAAACAAAT CTCTTATGGA GTTTATCTTT
5701 TTATCACTTG CAAAATATGT TTTTGTAAG AGATACTAAT TACTTAGTTA
5751 TTTGTAGTTA GCCATTCTTC TGATTAAAAA CCTAAAATTA AATCTTGAAA
5801 ATGTGTTTTC CTTCAAAACA CATCATTTGA GAGAAACACT AAAGTAAGTG
5851 TATGATTATC ATAGCATGTA CATAGGTGCT TCACAACCCA AAAAGAATAT
5901 TGTGATGGGT AAGAATCAGT AAAGGAATTT CTCCTAATAA AACAGTAGCC
5951 TATTAATTAA AGTAATGATA TGCAATACAG CAAGTTAAAG GGAAGTATC
6001 CTGGTGGGAT TATTGAAAGA TATACCCTTG ACTATAGATT AGAAAAACA
6051 GAGATGTTAT TTAGTGAAGA TATTGTGGTA CTCAATTATC ATCTGCAATT
6101 CACTTGCAGA GGAAAAATG AGTAATAAAT TCATTGTCAT TTTGGATTG
6151 TGTCTTTAAG TTGTGAAAAT AACTTAAAT ATAACCATCT GTCCTTTGCT
6201 CTTCTCTTCC TTTCTCTTT CTTCTCTTCC TTTCTCTTT CCATCCTTCC
6251 CTCCCTCCAT CTTCTCTTCC TTTCTCTTCT CTTCTCTTCC TTTCTTCTC

FIGURE 3, page 2 of 11


```

6301 TGTCTTCCT TCTTTTTTCC TTTCTTCTT TTTTCTTCTT TATTATTTCA
6351 TTAATCCCCC CTTCCATTTG ACGTCTAAAA GCCATGTTGT TCTAGAGGAC
6401 TTAAACTTAT TTTTCTCTTA ATAGCTTACT GAAAAATTAG TGATACAATT
6451 TTTTATTGTA ATTTGATGCT AATTCATTCT GTTATTTCTT TTATTGAGGA
6501 AGGCCCCACTA CATTATGTGA GACTGTGGGA AAAGCTGAGA TATGGCTAAT
6551 ACGAACATAT TGGGATTTTG AATTTCTCTA ACCATACCAA CCTAACTTTG
6601 AGTTTGTGTT AGGATTGCAC TGTAAACCTG CCAAAGCTTT GCCTAAGGTA
6651 GGACTATTGT ATTAAGGAAT ATTATGTACT TTATGACATG ACTTGTTTTT
6701 CCTTGAAAAGA TTACAACCTT AGTTATAGAA GGATGATGTT GAATGTCGTC
6751 TGTTTGCGAG TCCATATTTA TTTTCCATGC CACAGGGGCT CTTATAGGTG
6801 ATTATATGTC TTTTCGGTAT TATATGAGA AAGTAGGCAG AAGAATTTCA
6851 TGATTAGAAT AGATTTTAAA ATACTAGTAT TACAATAGTT TGGATAATAA
6901 ATTGAATTAA TAGGGAATTG GAGCCATGAA GATCACTAAA AAGAATGCTC
6951 TAGCCTTTCT CACAATCAAA TTGGGCTTAT GAACAAGGAT ATTTGTCATG
7001 ATAGTACAGA AATAAGCATA TTTTCATGAG ACATATTGGA TATATTCCAC
7051 AGGAGTTGGT GAGTGAGAGA AAATAAGTGA TGAAGGAAGA CAAAGAATAA
7101 AAGAAAAATT CAATAAATGG AAAGTTAAG TGTTTAATGA TAGTGATGAC
7151 TTTTACTCAA ATAAGTGCTT AGAAGTCATC TTGTTTGTGA TTTATATGAT
7201 GAATTCGTGT TTGTGACTAT CCACTTTGAG CTCGTGAGAA TGTTAGGTGA
7251 GGTTTAAATA AAGCCATTTG AGAAAAACAA GGTTTCAACC TCTGTGGACA
7301 GAAATCTAAA TATCGATAGT TATCAGGACA AAGTAGAGCT CATAGAAATA
7351 ATTTTGCAGC CTGCAGGTTT GTTTTGGAGT GAAAATAAAA TTGTATACTA
7401 TATTCCTAAA TCATCAGAGG AAAAAATTTA TAGTTCAAGG AATGTTGAAA
7451 GAAACAATAT TGAGAAGTAA AAGTGAGTAA TAGTTGTATAT AGTTTAAATA
7501 TAGTTTGTGA AGTATGTCTT GAGTTCCTG TCCCAAAAGT GGCTATTAGC
7551 TCTAGCCTTG ACCTGACAAG GTTCTAGGAT ATTTAGTCAT GGATGTTTCT
7601 AATCTACCTC TTACGGGATA CTTTTTATTC TGATGAACAG CTAATGCCT
7651 AAGTGTGCAA TCTATACCAA GATTGTCTT ATAGGGAAGT TGTTTACACT
7701 GGAAGACACC ACTGTGTCTC TTGTATGACC TATGTCTTCT TTATCCCTAC
7751 AAAGGTAACC ACATTATAGG AAACCTGAC AAGGCCAGAT GTTATATTTG
7801 TGTTGTGCAA GTGAGAAAAC ATGGGAGAAA CTTAACCAAA CACATAAAAT
7851 AACAGAAAACA GTCTTCTTTG ACCATTTCTA GAGAAAAGAG TTCAGCATCC
7901 CTTGTAAGGC CACTAGGAAG AAGAAAATTC TCTGGGAAAA GCACATTCAA
7951 CCAATGAATG GAGACCAAGA AAGAGAGTGA GGGATCTATG TGCCAAAATG
8001 TTAAGTGGGA TCCAGGGTGT TACCTAGGTG GGTTTCCAAT GGGGAACTGT
8051 AATTGGTAGG TTTAATGCAA GCAGGCACAA AGTCCATGGA GGCATTCTGA
8101 GACTGAAAGA TAGTCACTTT GGCATATCTG CACAGAATCT GATCAGTGAT
8151 TCAAGCCCAA GTAGGCTGTA TCTAGTTGTC CTATAGGGTG GTTACCAGGA
8201 GGCAGTGTGT AAGTAAAAAT CCTGACTGAA CACATTGAGG AAATGGAAGG
8251 AGGTGGAAGA TTTTAAACGG TGTCAGTGTT GACTAAGACC TGCTTCTGGT
8301 ATGGAAAATT CAACTTATAT TTTAAATGCA TAGCCAGACA ACATAAAATT
8351 ATAAGAATTT ACCACAATAG CTATGGTAAC AATACTGGGT TTACCTATTA
8401 CTACAGAGTG AAAAGAAAAC CCTCATTTCC CATTTTATGG AAATATAATC
8451 AAAATCCTAT AAGGAAGGTT TCAGAGCCAG TAGGATTTCC AGAAAAATTA
8501 TTGGTTTAT AGTAAGATGT GTATTGATGA ATATAATTTT ATTTATTAAT
8551 TATTAATATC ACTTTACTTA CCAGGAAAGT TATACCAGAA AACCAAGCTC
8601 TCTTAAGCCA TGGCATCTGT ATCTAAAATA GAAATACAGA AGGAGAGCTG
8651 ACAATTTCCA TCATTCTCTA GGTAATCTCC CATGCCATTC TACCCTTTAT
8701 TCCCACACTC CCAGTTTAC ACACACACAC AACACACAC ACACAAACAC
8751 ACACCATAG AAATAATCAT AGAAGACATA TTTTAAAAAA AGTTAGATCC
8801 ATACAGTAAT AATTTATTAG GTAAAAGCTT TTGTGCTGAT AATTTTACAA
8851 GTTTAATTGA GATATATTTT AGGGCTGTCT TACACTAAAT ATTTATTTT
8901 ATTTTTTAAA TTTGACATGT AATAATTGCA CATGTTTAA AGAAATGCTG
8951 TGGTATTACA ATACATTTAA ATGTTGTGTA ATAATTACAT CAAGATAATA
9001 AACCCATCAT CTAAATATTT ATCATTTCTT TGTGGTGATA ACATTCAAAA
9051 ACCTCCTTTC TGGCTATCTT GAAATATGTA ATACATTACT ATTAACATA
9101 GTTACCCAAC AACTTAATAT AATAACAGAA CATATTCTTC CAAATTTAAA
9151 CGTTGTATCC ATTGATCCAC CATTTCTCAT TGCCCTCCCT ACTATCTCTT
9201 CAGCCTCTAG TAACCACAAT TCTACTCTCT AATTATATTA TGAATGCATT
9251 TTTTGATTCC ACATATAAGG GATACCATGC TATCTCTGCC TGGATTATTT
9301 CAGTTAACAT TATGCCCTGG AGGTTTCATC ATGTTTCTAC AAATGACAGG
9351 ATTTCATCTT TTTTTTCCA ATATATATTT AATGAAATGG ATATATATAA
9401 ACATTGGAAA ATGTATATAT ATATATATAT CTCCAGTGGA ATGCTATTGA

```

FIGURE 3, page 3 of 11

9451 GCTATAAAAA AGTTAATATA TAATAGAAAT AAAGCTTATA TATATCTAAT
 9501 GGAATGGATA TATATATATA ATGGAATAGA AATATATATC TATACATATA
 9551 AACACACGCA ATATACATAT CCATTTCATT GCATATATAT ATATATAGAG
 9601 AGAGAGAGAG AGAGATATTT TCAAAATGTGT GTATATATAT CCAATGGAAT
 9651 GGACATATAT ATATGTATAT TTTTTCCTAT TTTTCTTTAT GTATTTCTTC
 9701 ATTAATGGAT GTTTAGGTTG ATTCATCCCT TGGGTATATG AATAATGTTG
 9751 ATGTAAACAT AGAAGGACAG ATATCTCTAT GACTTCTTAG TTTATTTAAA
 9801 TATACACCCA GTAATGGAAA TGCTGTATAA TATGGTAGTT CTATTTTCAT
 9851 TTTTGTAGGA ACTACCATAC CGTTTTCTCT ACTAATTGTA CTAATTTGCA
 9901 TTTCCCTCAA CAGTTTATAA AAGATCTTCT TTCTCTGCAT ACTTTCTAGC
 9951 ACTTGTTATT TTTGCCTTTT GATAATAGCC ATAACAGGGG TGATGTGATA
 10001 TCTCATTGTA GTTTTGATTT GCATTTCCTT GATGATTAGT GATTTTGAGC
 10051 ATTTTGTAAT TATACTTCTT AGTCACTGAT AGTCTTCTTT TGAGAAGTGT
 10101 CTATTCAGGT CTTTGTCTTA TTTTAAATC AAATTAGTAA TTTATTTTAA
 10151 TTGACTGATG TGACTTCTAT GTATATTTGA GATAGTAACT TATTGTCAGA
 10201 TTCATAGTTT GCAAAATATTT TTCATGTTGT GAATTGTCTC TTCACCTGT
 10251 TGTTTGCTTC ATTTTCTCTG CACAAGCTCA ATGCTTTGAT ATAACCCATT
 10301 TATCTACTTT TCCTTTTGTG GGCTGTGCTT CTGAAGTCCT ATCCAAAAAA
 10351 ATCCTTGCCCT AGACCAATGT CACAAATCAT TCCTCCTACA GTTTCTTCTA
 10401 GTAGTTGTAT AATGTTTGGC CTTATATTTA ACTTTGTAAT TCATTTTAC
 10451 TTACTTTGTA TATGGTGAGG GATAGAGGTC TAGTTTCATT TTCTGCATGT
 10501 GGATATGCGA TTTTCCTAGC ACCATTTAGT GAAGAGGTTG CCTTTTTTCT
 10551 ATTATGTGTT CTGGCACCT TTGTCAAAG TCAGTTAGCT GCTATATTC
 10601 TCCATTTGTG TTGTTATAGA GGAACACATG AGACTAGCAA ATTTATATAT
 10651 CAAATAGAAT TATTTGAATG ATAGTTCTGC ATACTGTACA AGAAGCACAG
 10701 CACTGACTTC TGCTTGGCCT CTGGTAAGGT TCTCAAGATG CTTCCACTTG
 10751 TGGTAGAAGG CAAACATGAG CTGGTATATG CAAAGGTCTC ATGACAAGAG
 10801 AGGAAACCAT AAAGAGGGGA TGTGAGGGAG TGCCAGGTTT TGTAAAACAA
 10851 CTAGCTCTTC TGGGAACATA TAGAGTAAAA ATTCGCCTCC CAGGCAGGGG
 10901 ATTAATCTAT TCATGAGGGA TCTGCTTCCA TGACAAAGGC ACATTCTGTT
 10951 AGATTCTACC CCCAATATTG GGGATCAAAT TTTAACATGA AGTGTGGAGG
 11001 GCTCAAATAT CCATACTATG GCAGCAGTAA ATGCATAAAT TTATTTTGTG
 11051 GATCTCTATT CTATATAGTA TTGGTGTATG TATCTGTTT CATGCCACTG
 11101 CCATACTGTT TTGGTGATGA TATCTATGCT ATATATGTGT GTGTGTATAT
 11151 ATATATTATA TATATGTATA TATGTGTATA TTATATATAT GTATATATGT
 11201 GTATATTATA TATATATAAT ACTTTAAGTT TTATATATAT ATAAAATACT
 11251 TTAAGTTCAA GGTACATGT GCAGGATGTG CAGGTCAGTT ACATAGGTAT
 11301 ACATGTGCCA TTTTGGTTTG CTGCATGCAT CAACTCATCA TTACATTAGG
 11351 TATTTCTCCT AATGCTATCC CTCCACCAGC CACCAACCC CCAACAGGCC
 11401 AGGTGTGTGA TGTCCCCGC CCTGTGTCCA TGTGTTCTCA TTGTTCACTT
 11451 CCTACCTAAA AGTGAGAACA TGCAGTGTTT GATTTTCTAT CCTTGTGATA
 11501 GTTTGTCTGAG AATGACTGTT TTCAGCTTCA TCCATGTCCC TCAAAAGGAC
 11551 ATGAACCTAT CTTATTTAT GGCTGCATAG TATTCCATGG TGTATATGTG
 11601 CTACGTTTTT TTAATCCAGT CTATCACTGT TGGACATTG GGTGGTTTCC
 11651 AAGTCTTTGC TATTGTGAAT AGTGCTACAA TAACCATATG TGTGCATGTG
 11701 TCTTTATAGC AACATGATTT ACTATCCTTT GTGTACATAC CCAGTAATGG
 11751 GATAACTGGG TCAAAATGGTA TTTCTAGTTC TAGATCCTTG AGGAATCCCC
 11801 ACACTGTCTT CCACAATGGT TGAACAAAT TACATTTCCA CCAACAGTGT
 11851 AAAAACGTTT CTATTTCCCC ACATCCTCTC CAGTATCTGT TGTTTCCTGA
 11901 CTTTTTAATG ATGGCCATTC TAACACACAT GAGATGGTAT CTCATTGTGG
 11951 TTTTGTGTTG CATTTCTCTG ATGACCAGTG ATGATGAGCA TTTTTCATG
 12001 TGTCTTTTGG CTGCATAAAT GTCTTCTTTT GACAAGTGTG TGTTCATATC
 12051 CTTTGCCAC TTTTCAATGG AGTTGTTTGT TTTTTCCTG TAAATTTGTT
 12101 TAAGTTCATT GTAGATTCTG GATATTAGCC CTTTGTGAGA TGGGTAGATT
 12151 GCAAAAATTT TCTCCCATTC TGTAGGTTGC CTGTTACCCC TGATGGTAGT
 12201 TTCTTTTGGT GTGCAGAAGC TCTTTAGCTT AATTAGATCC CATTTGTCAA
 12251 TTTCGGCTTT TGTGCCATT GCTTTTGGTG TTTTAGTCAT GAAACCCTTG
 12301 CCCAGGCCTA AGTCCTCAGT GGTATAGCCT AGGTTTTCTT CTAGGATTTT
 12351 TATGGTTTCA GGTCTAACAT TTAAGTCTTT AATCCATCTT AAATTAATTT
 12401 TTGTATAAGA TGTAGAAGG GATCCGTTT AACTTTCTAC ATATGGCTAG
 12451 CGTGTTTTCC CAACACCATT TATTAATAG GGAATCCTTT CTCCATTTCT
 12501 TGATTTTGTG ATATTTGTCA AACATCACAT GGTAGAGAT GTGTAGTGT
 12551 ATCACTGAGG CCTCTTTTCT GACTCCATTG ATCTATATAT CTGTTTTGAT

FIGURE 3, page 4 of 11

12601 ACCAATACCA TGTTGTTTTT GTTACTGCAA CCTTGTAATG CAATTTGACA
12651 TTCAGGACCA TGATGCCTCC AGTCTCTTTT TTTTCTCTA AATAATTTTT
12701 TTGTCAATGT AAGCTCATTT TCGTTCTTT CTGATCCATA AAGTATTTTT
12751 TTCCCATTTT GTGGAGAACG CCGCNNNNNN NNNNNNNNNN NNNNNNNNNN
12801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNGGCACA CCTCGTGCGC
12851 ATATATATAT ATATATATAT ATATACCTCT ATATATATAT ACATACATAC
12901 ATACATACAC ACCTCCTTGT CTGGTGTGGG ATCAGGGTAA TGCTAGCCTC
12951 ACAAGATGAT ACTGAAGTGT TTTTGCCTTT TTGACTTTTT GATGGTTTTG
13001 AAGAGTGAGA AAAAGTGTTA TTAATTATTC TTTAAATTTT GTTGAATTTT
13051 ATAGTGAAGA CCTTAGCTCA CTGGCTTTTT TAATGAGAAC TTTATTACTG
13101 ATTTAAACTT CTTCTTCATT ATTTATTTCT GCCTTGTTTT TATTTCTTCA
13151 TAATCCAGTC CTATTTTATG TGTCCACTAA ATTGTTTATT TTCCTAGAAT
13201 TTTTCCATTT ATTGGCATAT GCATGTCCAT AGAAGCCTTT TATAGTCCTT
13251 TTCATTCTA GTGTCATTTT TTTCTCTTTT TTTAAGAATC CTTAAGATTT
13301 TAGAGATGAA ATGTCACCTT GTTACGCATA CTGGAGTGCG GTGACATTAT
13351 TATAGCTCAC TGAAACCCAA ACTCCTGAG TTAAGCAATC CTTCTACCTC
13401 AAAATTCCAA AATTCCTGAG TAGCTGAGAC AGGCATACAC CATCAAGACT
13451 GGCTAATTTA TTTCAAATTT TGTAGAGATG GGTCTTACT AAGCTATTCT
13501 CAATCTTTGG GCTTCAAGTG ATTCTTCAGC CTCTGTCTCT GAAAATGCTG
13551 GGTTTATAGA TATGAGCCTC TATGCCTGAT TTGCTTTGTC TCTTTGTAAT
13601 CTCCCATTTT ATTTGTGTCT TTTCTGGTTT GTTTCATTTT GTTATGTTTT
13651 CAGTTACCTT GCTAAAGCTT TGTCGATTTT ATCTCTTCAA ACAACTAACT
13701 CAATATTTTG CTGATTTTCC ATATAGTATT TTATTCTAT TTCATTTATT
13751 TCTGCTCTAA TCTTTGTAA ATATCTTGT TCTCTAATAA TTTTGAGTTT
13801 CCTTGTTCTT GTTTTCTAAT TCCTTGCGAT GTTATCATAA ATTGTTTATT
13851 TGATATCTTT CTACTTTTTT GATGTGTGTG TTCGTGTTG TAGACTTTCC
13901 TCTTTATAT TCTGATTTCT TCCTCAATTC TCTAATATTA TGATTGCATT
13951 ATTTTCCAAG TTTCTTTTGT TTTTATTTT ATAGTTTATG TGATTCTCTGA
14001 ACTTGTCAAA GAGATTATTG TGAATTTGAT GTCGGATATT TAAGCATTTT
14051 CAAAACCTTT GGTGCATTAT TGAAATTTA TTGGTTTATT TTAGAGATGT
14101 CATACTTCCC AGTTTTTTTT TAACAATACT TGCTCTTTAT ATTGATGTCT
14151 ACATATTTAA AAAGATAACC ACCTGATTCA GCTTTTAAAG GTGATATGCA
14201 GTGGTGTTAA GTGTGTACTG CTTAATATCA GAGCTGAATC ACTGCCCTGA
14251 GGATTCCTTC TGTTCTGAGG AGAGCTTGTA GTTAATAGCA GAACCTAAAT
14301 AGTGCAGTAG AGCTAAATCT CTTCCATGCT GTTGTTTTCC TGCTGGGGA
14351 AGACTTATCA TGACCATGAA AACATAATGC TGTGCCAGAA CTTAAACCCA
14401 AACCTGTAGT AATTCTGAG TTGAGGAAG CTTAAGAAAT AACTGGAAC
14451 TAGTTACTAA CCTGATAGTT GTTTCTGAGT CAGAGAAATG CTCTGCATGA
14501 TCACCTGGGA TATTTGTAAA ATCTAACCAA AGATTCTAGC CTTCCCCTTG
14551 GATTGTGCCT CCTGTACTAC TGTAGTGCTG GCTAGGTCCT CATCAGTGAA
14601 TTCCCTGCTG ATAGACCAC AAAGCATCTG CCAAGATCTG TTTGCCATTT
14651 GCTGTGATTA GTGCTTCTGC TCTTTGCTTC CAATTCAACT CAGGTGGTTC
14701 AGCCCTTCTG ACATCCTAA TACCTCCTGT GGGATGGAAC ATAGAAGGCT
14751 TCTCACAATG ATTCACACAC TGATATGGAG ATTGAATGTC CAGTTGCAAC
14801 TATTTTCTTC CACCTGTGTA ATTGCAGGTA CAGGGAAGTT TTCTGTGACT
14851 GATGCTATTT TGGTTTGGAG AATGGGGTGA TGTGGCACA TGATCTTTCT
14901 TCTTTCTGGT CATGGATTTT TTAATTCCA TGAACCCATA AGATTTTCA
14951 CTTTCTTCT GAGCTCTGGT GCTTTCAGAG TGGTATTTTT ATATTGCAAT
15001 AGTTGCTAGT TGTACTTTTA AAAGCGATTG ATGCTGGAGG TCTTCTATTC
15051 CACCATTCCG CTGATGTCAG TCCTCAAATA ATAATTTTAT ATTTTAGCAA
15101 ATTATTTTGG TTTTAGGATT TTGTGTCTAC GTGACACAGA CATGAAAAGA
15151 GATGTACTCA TTAGTGAAAC TTTTGCATA CTGTTTGGT TGTGCGCCTT
15201 TTCTAGTATG AATGATTACA TATTTAAGCC ACATGTTTTA TACATAGACT
15251 GTCCTTTAAA GAGACTAGAT AGTTCTGTGT GTCAGCATAT AGGGACAGAA
15301 TATAACTACA CATTAATAAT TTCTCAAGTA TTTATTTTAG AAGTGTAAGT
15351 AACCTTTTAT TTAATTTTGG TTATATTATG CCTCTGTAAT GCAGATAAAT
15401 TTTTATCTTC AGGAAATGGA AAATTTGTG CAGAGTTTCA GGAAGATGG
15451 TATTGTGGTG TTTTCTCTGG GGTCACGTG TCAAAATGTT ACAGAAGAAA
15501 AGGCTAATAT CATTGCTTCA GCCCTTGCCC AGATCCACA GAAGGTCAGT
15551 AAAACCTCCA ATCCTGATAA GCAGCTATTC ACATAATGAA ACAGTATGGT
15601 TTTATTTGGG TCTTGAATCT CATTTTCCAC TTAGCATAAC AGGTACCAA
15651 ATTTGCAAAA CATTATAGTA GTGTACATGG GCATAACTGA TCATTTGCCT
15701 ACTGAGTCTT GCTGTTACTG GAAACAACCT TCTTGATTGT CATTTGTTTA

FIGURE 3, page 5 of 11

15751 TAATAAAATA GATATAATAA ATAAAGCTCT ACCTTATATT TTAGGATTTG
15801 AAATCTAAAA GCGTGTGCCA ATGATTCCAA AAAAAAATTC TGACATCTAT
15851 TATTTCAAAG GACCAGAAAA AGGAAAACTG ATATAAAAAA AAAAAAGAAGA
15901 ATCAATCTCA AGAATATCTT CTCATATTTG TGTGTATAAA AACTGTATTG
15951 AGGGTAGTTT TGCTTAGAAA TAAAAGCTCA GATTAATGTA GTCTTTCTAA
16001 ATAATTAGAA GTTTCAAAAG TAAAATGTCA ATTACAATTA TAGTATAGTA
16051 ACAATTATTT AAGTAATGTA ATTATTTATG ATACTCCACT AATTTTAACT
16101 TTATTATTAC TGTAATTCTA GAATTTCACT CTTTAGATAG TGCTATATAT
16151 AAACATATCA AAAGATATTT CATTTTATAT TTAGCTAAAA TACTTCAAAC
16201 TCAATAAAGG CAAGCATACT AATTAGGAAT TTGAAATATT GTAATTTCAA
16251 TTATGAAAAT ATCTGTTAAG TAGTTTGAAA CATCTATGCC GTTCTTTGTT
16301 TTCAAATGTA TAAAATTTGT ATAGGTGTCC AACAAAGAAA AATTGTGTAA
16351 AAAAAAGGTA CAATCTCAA GAAAATTTAT CATTGAACAG TGGAACATAA
16401 GTAATTTTCT AGCTCATTCT TCTTCAATAA AACAATTAAA TATAAGAAGA
16451 AAGAGGCCAG GAAGGAAATA GAGAAGAAA GACCCCGAT TATCCAAAAG
16501 ACACACATAA TTGAAAGCAA ATTTTATCT GCAGGGAAC GTAAATTTGA
16551 TGGTAGAATG AGATTGGCTC CATGAGTTAA AATGACACAG AGATCAGGTA
16601 CTTATGAAAAT TTTTAAATCT TATATAAAAA TAGATTAGCC ACTGCTGAAT
16651 TATTTTTTTA AATATTCAC GGTATTCTCA TTCTCAAATA TTTTAAATTG
16701 GTAATAAAAT AATAATAGCA TACCTAATAG GCAACTGGTA CACATTATTT
16751 TAAAAGATCT TTGTAAACG TCCTACTATA TCTTTCAGTC TTTACGCGGT
16801 AGCTCTACAC ACCCTGTCT CAACCATCAC CTGAAGTACA ATGAGTTTAT
16851 AATTTATAAC TATATCTACA TCCTTAGAAT GCTAATATCC TGTGGTTCAC
16901 TCTGTGAAAT ACATGTGTTT CTTCCTAGG TGTATGGAG GTACAAAGGA
16951 AAAAAACCAT CCACATTAGG AGCCAATACT CGGNNNNNNN NNNNNNNNN
17001 NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNNNNNNN NNNAAAAAA
17051 AAGACCAAT CCCAAAGAAA ATTTATCATA GAACAATGGA ACATAGGTAA
17101 TTCTCTAGCT CATTCGTCTT CAATAAAACA AATAAATATA AGAAGACAAA
17151 GGTCAGGAAG GAAATAGAGA AGAAAAGATA ACCGATTATC CAAAATCACA
17201 CACAAAATTG AAAGCAAATT TTATCTGTGG GGAAGTGTAA ATTTGATGGT
17251 AGAACCAGAA TAGTTCCATG ATTTGAAATG ACACAGAGAT CATGTACTTA
17301 TAAAATATTT TATTCTTATA AGAAAATTGA GTAGCCAGTG CTGAATTACT
17351 TTTTAAATGAT TCACTGATAT TCTCATACTC AGATATTTTA ATTGATATTA
17401 AAATAAATAT AGTATACTTA ATAGTCAACT GGTACACATT ATTTGAAAGG
17451 ACTTTTGATA AAAGTCTTAC TATGTCTTTT ACTGTTTACA CAGTACCTCT
17501 ACATACCCCT GTCTCAACCA ACACCTGAAG TACAATGAGT TTATAATTTA
17551 TAACTATATC TACATCCTTA GAGTGCTAAT ATCCTGTGGT TCAATCTGTG
17601 AAATACATGT GTTCTTCCA TAGGTGTTAT AGAGATACAA TGGAAAAAAA
17651 ACCATCCACA TTAGGAACCA ATACTCGGCT GTATGATGGG ATACCCAGAG
17701 ATGATCTTCT TGGTAGGTCT ATGAGAAAGT AAAAATATGA ACTAGACGAG
17751 GAAAAAATGA ATAAATGTTA AACAGCAAGC AAATTCAGCA AAGATCTAAA
17801 ATTATAAAAC TTTATTTTAC TTACTCTTTT GAAGCAGATA TAATTAAAGG
17851 ATTGACTAAA ATTGATAGTA TTCACACTTT CTATTGTAA GGTGAGAGTG
17901 ACAGGAAAT CAGAAAGGAT TAATGCCTAT TTTTCTGGAG ATAGAAATGA
17951 TCTTTAGTAG CAATGCTCCA TGTGCTCACC TTCTAAAGAA AGTGCTGTAC
18001 GCTTCAGTGA GTTATCTCGT AATTCCCATC TGTAGTTTTT AAATAATTTT
18051 AAAAGTTTAG AATAAAATAT CTCACCATT CTATCCAAT TTACATACTA
18101 GGTCATCCCA AAACCAAAGC TTTTATCACT CATGGTGGAA TGAATGGGAT
18151 CTATGAAGCT ATTTACCATG GGGTCCCTAT GGTGGGAGTT CCCATATTTG
18201 GTGATCAGCT TGATAACATA GCTCACATGA AGGCCAAAGG AGCAGCTGTA
18251 GAAATAAACT TCAAAACTAT GACAAGCGAA GATTACTGA GGGCTTTGAG
18301 AACAGTCATT ACCGATTCCT CGTAAGTACT ACTGCTTGTA CAGACTGATC
18351 TAACATTGAC TATGTTATAC ATTATACCAG AAAATGTTAA ATATCATCCT
18401 GGTAGACATG TTGAGGGATT TTACTCCACA ATATTGAGTC ATTCATCACC
18451 TTGTTACTGG AATAGTTGTG GAAATTGTAG TTCATAGAGT GTCAAACCTT
18501 CTTCATGGAA ATATTAGGTT TAAGTTAACT ACTGGCTTAC TAAGCTTTTA
18551 TTCACATCTT AATTTTACCC CATTTTGTTA AGAATATACT CTTTCAGTCT
18601 CTCCACTATA TCTGTTTAACT ACTATGTAAC CAACAATATT CATGTCACAA
18651 CCAGAATCAA TCTTTTACTG AACATGTTCT TGGCTTGCAT AACATATACT
18701 ACGGTTTATC TACGTGTCTT TTATGAAAAC AAAACTACAA CTTTCTAAGT
18751 TCTATGTGTG TTTTCCCTT CCAGTTATAA AGAGAATGCT ATGAGATTAT
18801 CAAGAATTCA CCATGATCAA CCTGTAAAGC CCCTAGATCG AGCAGTCTTC
18851 TGGATCGAGT TTGTCATGCG CCACAAAGGA GCCAAGCACC TCGCATCAGC

FIGURE 3, page 6 of 11

ALLELIC VARIANTS (SNPs) :

DNA

Position	Major	Minor	Domain
1735	A	G	Intron
1922	A	G	Intron
2361	C	T	Intron
7371	G	C	Intron
9558	G	A	Intron
10579	T	G A	Intron
10625	C	T	Intron
11147	A	G	Intron
15131	C	T G	Intron
15221	A	G T	Intron
15778	T	C	Intron
15895	-	A	Intron
19786	-	T	Beyond ORF (3')
20157	G	A	Beyond ORF (3')
20246	T	C	Beyond ORF (3')
20681	C	A	Beyond ORF (3')
20819	T	C	Beyond ORF (3')

Context:

DNA

Position

1735	TACTTTACTGATATACACCTATTGACTTATTTTATTTCTTTGCTTTACTGATAAAAAG TTGTTTTGCTTTGCAATTTTCATATAGTTGTGATCAGAGCTGGTCAATGCAAGACATGTT TTTATCCAAATATGTTTGAGAAATATGTAGAAACATGAAAAAGGTACAATTATATCCGA CACTAAAATATTGTTTAATGTATTTCCAACGAATCTTATGCATAGACTGTTTCACAGAAC TAATATTTCAGAGGATCCCAGTTCAAATGTCCTTAGCCTTAGACATGATTTGAATTTACAT [A, G] TATTGATTTGCTTTAAATAATTTTCCATTTCAGTAAGCTGTGCCTAGCTGCAGATAGCCTA CCAGGCTTTATGGATCTAGGTAAACAATACAAATCTCTTGGCCTCAAGTCTACATTTCAGA TATTAATTTAAAGGGGTACAGCTATATAGAGGTCACTGGCAAATTTTGGTAAATAGGAT TATAGTAAAAGCCCCCTGACAAGATTGAAATTTAAATAAAAACAAAAGTGTATCAAAGG GGTGAAAGAGCATTTTCCAATAAACAAAAGTGGGTTCTGGCCATGCATTTCAGAAATTTCC
1922	ATATTGTTTAATGTATTCCAACGAATTTCTTATGCATAGACTGTTTCACAGAACTAATATT CAGAGGATCCCAGTTCAAATGTCCTTAGCCTTAGACATGATTTGAATTTACATATATTGA TTTGCTTTAAATAATTTTCCATTTCAGTAAGCTGTGCCTAGCTGCAGATAGCCTACCAGGC TTTATGGATCTAGGTAAACAATACAAATCTCTTGGCCTCAAGTCTACATTTCAGATATTAA TTTAAAGGGGTACAGCTATATAGAGGTCACTGGCAAATTTTGGTAAATAGGATTATAGT [A, G] AAAGCCCCCTGACAAGATTGAAATTTAAATAAAAACAAAAGTGTATCAAAGGGGTGAAA GAGCATTTTCCAATAAACAAAAGTGGGTTCTGGCCATGCATTTCAGAAATTTCCCAACAAT TCTTTAAAATCATGGAGCAGCTTGATATATAAGAAATTCATTTAATAACTATATTTATT ATGTAGCTCCAACTTACTAAATTATTGATTATTATATATTTTATAGAATTATCTATTGTG AGTCTAAATCAAGAGTATATATTCAAACAACATATAGGAAAAGGGATATCAGTCAATTTCA
2361	CAGCTTGATATATAAGAAATTCATTTAATAACTATATTTATTATGTAGCTCCAACTTACT AAATTATTGATTATTATATATTTTATAGAATTATCTATTGTGAGTCTAAATCAAGAGTAT ATATTCAAACAACATATAGGAAAAGGGATATCAGTCAATTTCAATTCAAGGATTTATTTCC ATAAGTGCTTACGCACAGGTGTATTTCAATTTTATTATACATTGCTTTATTGTCCTTCACA AAAATTGCAATTTACAAATTAAAGGTTTTTGAAAACCTTGAATCAAGCTAATCAATTTGG [C, T] GTAATATTTCCAACAACAAGTGTGACTTTTGGACTCTATCACATATTGGCATTATCATG CTTTTTCAAATTTTTCATTGTTATATCTGTTACGGTGATCTGGGATCAGTGTTCCTTGAT GGTTACACGTTTATTAGCTTGGGGGCACCTTGATGTGTTACAATATAAGACAGCAAACCTT AATTATAAATGTTGTGCATGTACTAACTGCTCCGCTGATTCGTTTCCCCATCCCACTTCT TCTTAGGCCTCCCTATTCCCTGAGACACAGTAATATAACATACAATGACTTCTAAATGTT

7371 AAATAAGTGATGAAGGAAGACAAAGAATAAAAGAAAATTTCAATAAATGGAAAGTTTAAAG
TGTTTAATGATAGTGATGACTTTTACTCAAATAAGTGCTTAGAAGTCATCTTGTGTGA
TTTATATGATGAATTCTGTGTGTGACTATCCACTTTGAGCTCGTGAGAATGTTAGGTGA
GGTTTAATAAAAGCCATTTGAGAAAAACAAGGTTTCAACCTCTGTGGACAGAAATCTAAA
TATCGATAGTTATCAGGACAAAGTAGAGCTCATAGAAATAATTTGCAGCCTGCAGGTTT
[G, C]
TTTTGGAGTGAAAATAAAATTGTATACTATATTCCTAAATCATCAGAGGAAAAAATTTAT
AGTTCAAGGAATGTTGAAAGAAACAATATTGAGAAGTAAAAGTGAGTAATAGTTGTTATA
GTTTTTTAATAGTTTGTAAAGTATGTCTTGAGTTCACTGTCCCAAAGTGGCTATTAGCT
CTAGCCTTGACCTGACAAGGTTCTAGGATATTTAGTCATGGATGTTTATAATCTACCTCT
TACGGGATACTTTTTATTCTGATGAACAGCCTAATGCCTAAGTGTGCAATCTATACCAAG

9558 TCCACATATAAGGGATACCATGCTATCTCTGCCTGGATTATTTAGTTAACATTATGCCC
TGGAGGTTTCATTCATGTTTCTACAAATGACAGGATTTTCATTCTTTTTTTCCAATATATA
TTTAAATGAAATGGATATATATAAAACATTGGAAAATGTATATATATATATATCTCCAGT
GGAATGCTATTGAGCTATAAAAAAGTTAATATATAATAGAAAATAAGCTTATATATATCT
AATGGAATGGATATATATATATAATGGAATAGAAATATATATCTATACATATAAACACAC
[G, A]
CAATATACATATCCATTTTCATTGCATATATATATATATAGAGAGAGAGAGAGAGATAT
TTTCAAATGTGTGTATATATATCCAATGGAATGGACATATATATATGTATATTTTTTCCA
TATTTTCTTTATGTATTTCTTCATTAATGGATGTTTAGGTTGATTTCATCCCTTGGGTATA
TGAATAATGTTGATGTAAACATAGAAGGACAGATATCTCTATGACTTCTTAGTTTATTTA
AATATACACCCAGTAATGGAATGCTGTATAATATGGTAGTTCTATTTTCATTTTTTGAG

10579 CAATGCTTTGATATAACCCATTTATCTACTTTTCTTTTGTGGCTGTGCTTCTGAAGTC
CTATCCAAAAAATCCTTGCCCTAGACCAATGTCACAAATCATTCCTCCTACAGTTTCTTC
TAGTAGTTGTATAATGTTTGGCCTTATATTTAACTTTGTAATTCATTTTTACTTACTTTG
TATATGGTGAGGGATAGAGGTCAGTTTCATTTTCTGCATGTGGATATGCAGTTTTCTTA
GCACCATTTAGTGAAGAGGTTGCCTTTTTTCTATTATGTGTTCTTGGCACCTTTGTCAA
[T, G, A]
GTCAGTTAGCTGCTATATTCCTCCATTTGTGTTGTTATAGAGGAACACATGAGACTAGCA
AATTTATATATCAAATAGAATTATTTGAATGATAGTTCTGCATACTGTACAAGAAGCACA
GCACTGACTTCTGCTTGGCCTCTGGTAAGGTTCTCAAGATGCTTCCACTTGTGGTAGAAG
GCAAACATGAGCTGGTATATGCAAAGGTCCTCATGACAAGAGAGGAAACCATAAAGAGGGG
ATGTGAGGGAGTGCCAGGTTTTGTAAAACAACAGCTCTTCTGGGAACATAAGAGTAA

10625 GTGCTTCTGAAGTCCTATCCAAAAAATCCTTGCCCTAGACCAATGTCACAAATCATTCCT
CCTACAGTTTCTTCTAGTAGTTGTATAATGTTTGGCCTTATATTTAACTTTGTAATTCAT
TTTTACTTACTTTGTATATGGTGAGGGATAGAGGCTAGTTTCATTTTCTGCATGTGGAT
ATGCAGTTTTCTAGCACCATTAGTGAAGAGGTTGCCTTTTTTCTATTATGTGTTCTTG
GCACCTTTGTCAAAGTCAGTTAGCTGCTATATTCCTCCATTTGTGTTGTTATAGAGGAA
[C, T]
ACATGAGACTAGCAAATTTATATATCAAATAGAATTATTTGAATGATAGTTCTGCATACT
GTACAAGAAGCACAGCACTGACTTCTGCTTGGCCTCTGGTAAGGTTCTCAAGATGCTTCC
ACTTGTGGTAGAAGGCAAACATGAGCTGGTATATGCAAAGGTCCTCATGACAAGAGAGGAA
ACCATAAAGAGGGGATGTGAGGGAGTGCCAGGTTTTGTAAAACAACAGCTCTTCTGGGA
ACTAATAGAGTAAAAATTCGCCTCCAGGCAGGGGATTAATCTATTTCATGAGGGATCTGC

11147 ACAACTAGCTCTTCTGGGAACATAAGAGTAAAAATTCGCCTCCAGGCAGGGGATTAAT
CTATTCATGAGGGATCTGCTTCCATGACAAAGGCACATTCTGTAGATTCTACCCCCAAT
ATTGGGGATCAAATTTTAACATGAAGTGTGGAGGGCTCAAATATCCATACTATGGCAGCA
GTAAATGCATAAAATTTATTTTGTGGATCTCTATCTATATAGTATTGGTGTATGTATCTG
TTTTTCATGCCACTGCCATACTGTTTTGGTGATGATATCTATGCTATATATGTGTGTGT
[A, G]
TATATATATTATATATATGTATATATGTGTATATTATATATATGTATATATGTGTATATT
ATATATATATAAATACTTTAAGTTTATATATATATAAAATACTTTAAGTTCAAGGGTACA
TGTGCAGGATGTGCAGGTCAAGTACATAGGTATACATGTGCCATTTTGGTTTGTGCTGCATG
CATCAACTCATCATTACATTAGGTATTTCTCCTAATGCTATCCCTCCACCAGCCACCCAA
CCCCAACAGGCCAGGTGTGTGATGTTCCCCGCCCTGTGTCCATGTGTTCTCATTGTTCA

15131 CAGGGAAGTTTTCTGTGACTGATGCTATTTTGGTTTGGAGAATGGGGTGATGTGGCACAA
TGATCTTTCTTTCTTCTGGTCATGGATTTTTTAATTTCCATGAACCCATAAGATTTTTCA
CTTTTCTTCTGAGCTCTGGTGCTTTTCAGAGTGGTATTTTTATATTCTGAATAGTTGCTAGT

FIGURE 3, page 9 of 11

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TGTACTTTTAAAAGCGATTGATGCTGGAGGTCTTCTATTCCACCATCTCGCTGATGTCAG
TCCTCAAATAATAATTTTATATTTTAGCAAATTATTTGGTTTTAGGATTTTGTGTCTAC
[C, T, G]
TGACACAGACATGAAAAGAGATGTACTCATTACTGAACTTTTTGCATACTGTTTTGGTT
GTGCGCCTTTTCTAGTATGAATGATTACATATTTAAGCCACATGTTTTATACATAGACTG
TCCTTTAAAGAGACTAGATAGTTCTGTGTGTGTCAGCATATAGGGACAGAATAAATACTACAC
ATTAATAATTTCTCAAGTATTTATTTTAGAAGTGTAAAGTAACCTTTATTTAATTTTTGT
TATATTATGCCTCTGTAATGCAGATAAATTTTTATCTTCAGGAAATGGAAAATTTTGTCC
15221 TTAATTTCCATGAACCCATAAGATTTTTCACTTTTCTTCTGAGCTCTGGTGTTCAGAG
TGGTATTTTTATATTCGAATAGTTGCTAGTTGTACTTTTAAAAGCGATTGATGCTGGAGG
TCTTCTATTCCACCATCTCGCTGATGTGAGTCTCAAATAATAATTTTATATTTTAGCAA
ATTATTTTGGTTTTAGGATTTTGTGTCTACGTGACACAGACATGAAAAGAGATGTACTCA
TTACTGAACTTTTTGCATACTGTTTTGGTTGTGCGCCTTTTCTAGTATGAATGATTACA
[A, G, T]
ATTTAAGCCACATGTTTTATACATAGACTGTCCTTTAAAGAGACTAGATAGTTCTGTGTG
TCAGCATATAGGGACAGAATAAATACTACACATTAATAATTTCTCAAGTATTTATTTTAGA
AGTGTAAAGTAACCTTTATTTAATTTTTGTTATATTATGCCTCTGTAATGCAGATAAATT
TTTATCTTCAGGAAATGGAAAATTTGTCCAGAGTTCAGGGGAAGATGGTATTGTGGTGT
TTTCTCTGGGGTCACTGTTTCAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAG
15778 GTTTCAAAATGTTACAGAAGAAAAGGCTAATATCATTGCTTCAGCCCTTGCCCAGATCCC
ACAGAAGTCTAGTAAAACCTCCAATCCTGATAAGCAGCTATTACATAATGAAACAGTAT
GGTTTTATTTGGGTCTTGAATCTCATTTTCCACTTAGCATAACAGGTACCAAATTTGCA
AAACATTATAGTAGTGTACATGGGCATAACTGATCATTGCTTACTGAGTCTTGCTGTGA
CTGGAAACAACCTTTCTTGATTGTCAATTTGTTTATAATAAAATAGATATAATAAATAAGC
[T, C]
CTACCTTATATTTTAGGATTTGAAATCTAAAAGCGTGTGCCAATGATTCCAAAAAAT
TCTGACATCTATTATTTCAAAGGACCAGAAAAAGGAAAACCTGATATAAAAAAAGAA
GAATCAATCTCAAGAATATCTTCTCATATTTGTGTGTATAAAAACTGTATTCAGGGTAGT
TTTGCTTAGAAAATAAAAGCTCAGATTAATGTAGTCTTTCTAAATAATTAGAAGTTTCAA
AGTAAATGTCAATTACAATTATAGTATAGTAACAATTATTTAAGTAATGTAATTATTTA
15895 TATGGTTTTATTTGGGTCTTGAATCTCATTTTCCACTTAGCATAACAGGTACCAAATTT
GCAAAACATTATAGTAGTGTACATGGGCATAACTGATCATTGCTTACTGAGTCTTGCTG
TTACTGGAAACAACCTTTCTTGATTGTCAATTTGTTTATAATAAAATAGATATAATAAATA
AGCTCACTCTTATATTTAGGATTTGAAATCTAAAAGCGTGTGCCAATGATTCCAAAGAA
AAATCTGACATCTATTATTTCAAAGGACCAGAAAAAGGAAAACCTGATATAAAAAA
[-, A]
GAAGAATCAATCTCAAGAATATCTTCTCATATTTGTGTGTATAAAAACTGTATTCAGGGT
AGTTTTGCTTAGAAATAAAAGCTCAGATTAATGTAGTCTTTCTAAATAATTAGAAGTTTC
AAAAGTAAATGTCAATTACAATTATAGTATAGTAACAATTATTTAAGTAATGTAATTAT
TTATGATACTCCACTAATTTTAACTTTATTTACTGTAATTCTAGAATTCACACTTTA
GATAGTGTATATATAAACTATCCAAAAGATATTTCAATTTTATATTTAGCTAAAATACTT
19786 GAAACATAATGATGTTTCTTTTCAATTTAAATAAGCCCTTCTACATAGCCAGCATCAGT
GATCTCAGAAAAATAATTGCTAATAATGATGACATGGCATTATGCTTAGAAAAAGTTTGT
GTATTTCCATAGACCTCATCTAGATGTCATGGCCTACATTTCTGCCATCACTCAACCAAT
ACTTTTTTCTGTTTTCTTGATGATAAAAGACCTTTCTCATGATTGCCATCAAATAACAA
AAGAACTATTTTTTTTCTCACATAGAGAACATGTCAGTAAGATATTCAAGGTGAACAGA
[-, T]
ATTTTTGGGATTAGTAACATTTGAAATATGTGGTGATAATTACTGAGTTTATAAAATTT
ATTTGATAGTACACTTAAAGAAGATTTATATGTTTATCTTTAAAAATGATGAATACTCA
TAATCTTATCTCTATAATCAAAAGTATAATTTACTGTAGAAAAATAAAGAGATGCTTGT
TCTGAAAGTAAGATCAGTGAAGTCTTTTCTAGTCTCAATCTTTGAGAATTGTAAATTCAT
CAAATAATTGCTTACATAGTAAAAATTTAAGGTATTAGAAAACCTGCATAACAAATAGTA
20157 ACACTTAAAGAAGATTTATATGTTTATTTCTTTAAAAATGATGAATACTCATAATTCTTAT
CTCTATAATCAAAAGTATAATTTACTGTAGAAAAATAAAGAGATGCTTGTCTGAAAGTA
AGATCAGTGAAGTCTTTTCTAGTCTCAATCTTTGAGAATTGTAAATTCATCAAATAATTG
CTTACATAGTAAAAATTTAAGGTATTAGAAAACCTGCATAACAAATAGTATTATATATTA
AATATTTTGTATGTAAAGCTCTACACAAAGCTAAATATAGTGTAAATATGTTTACACTA
[G, A]

FIGURE 3, page 10 of 11

TAAGCAAATATGTTAATCTTCTCATTTTTTTACTGTGTCATATAATCTTAGTGATATGCCTA
T'TAATAGTTTTAAATAAAATAAATTGGCTTATCTGGCTTTTTGAAAATTTTGAAATCTTA
CAGATGTTGATTAGGTATATCTACAAATTAATTTCAATTTTAAAATGATGATATAAAAAAT
AAATATAAGTATTTTTCTTGTGTATGTATACAATAAATAAATAAATTGTTTACTGTT
TTGAAAGTTTCTTAAGTTTTTACACTGATATGTTTTTTGACTTTTACAATATTATTATAA

20246 GAAAAATAAAGAGATGCTTGTTCTGAAAGTAAGATCAGTGAAGTCTTTTCAGTCTCAAT
CTTTGAGAATTGTAAATTCATCAAATAATTGCTTACATAGTAAAAATTTAAGGTATTAGA
AAACCTGCATAACAAATAGTATTATATATTAAATATTTTGATATGTAAAGCTCTACACAA
AGCTAAATATAGTGTAAATAATGTTTACACTAGTAAGCAAATATGTTAATCTTCTCATTTT
TTTACTGTGTCATATAATCTTAGTGATATGCCTATTAATAGTTTTTAAATAAATAAATTGGCT
[T, C]
ATCTGGCTTTTTGAAAATTTTGAAATTTCTTACAGATGTTGATTAGGTATATCTACAAATT
AATTTCAATTTTAAAATGATGATATAAAAAATAAATAAGTATTTTTCTTGTGTATGTAT
ACAATAAATAAATAAATAAATTGTTTACTGTTTTGAAAGTTTCTTAAGTTTTTACACTGAT
ATGTTTTTTGACTTTTACAATATTATTATAATCTAGGAAAAGCTGATTATATCTGTTTTA
AGCCTCATCTTTTCTCTGTAATTAAACACAGTAATTTATTAAACATGCTGTGACAGGTGGG

20681 TAAAATTGTTTACTGTTTTGAAAGTTTCTTAAGTTTTTACACTGATATGTTTTTTGACTT
TTACAATATTATTATAATCTAGGAAAAGCTGATTATATCTGTTTTAAGCCTCATCTTTTC
TCTGTAATTAACACAGTAATTTATTAACATGCTGTGACAGGTGGGAAGCCATTTCTGGA
GTTGAGCCTGCTGACACTCTGGAGCTTTTTAGGTTGGACGTTTATTGTATGTGGGACTCT
CTGCCTCTCGATAGCTGTTGCTCATAAGACTCTCCTTCATCAATCTGGCATTGAATTTTG
[C, A]
GATCAGTTGCAATCAGAATCCAATTGGCCTTGCCGTTTTAGTATGTTCTATCTTAACCAG
CAATTTCTAACCAGGAGCCTGCCAGGTTTGTCTGTCTTCCCTGTAAGAAGCTCCCAGC
ATAAATATTCTAAATTTTACACTACTAATCTATTAACCAACCTTTGGACCATGTTCACTT
TAGGTTGAGCATAGTGTGATGAGATGCAAATTAAATTACAATCCTATAGGTGTGTGTTAT
AAATTTTAAAGTGATATAAATTAAATAACACATTCTAAGTATCCAACAAAGGTCAAAAAA

20819 AATTTATTAACATGCTGTGACAGGTGGGAAGCCATTTCTGGAGTTGAGCCTGCTGACACT
CTGGAGCTTTTTAGGTTGGACGTTTATTGTATGTGGGACTCTCTGCCTCTCGATAGCTGT
TGCTCATAAGACTCTCCTTCATCAATCTGGCATTGAATTTTGGCATCAGTTGCAATCAGA
ATCCAATTGGCCTTGCCGTTTTAGTATGTTCTATCTTAACCAGCAATTTCTAACCAGGAG
CCTGCCAGGTTTGTCTGTCTTCCCTGTAAGAAGCTCCCAGCATAAATATTCTAAATTT
[T, C]
ACACTACTAATCTATTAACCAACCTTTGGACCATGTTTCACTTTAGGTTGAGCATAGTGTG
ATGAGATGCAAATTAAATTACAATCCTATAGGTGTGTGTTATAAATTTTAAAGTGATATA
ATTAAATAACACATTCTAAGTATCCAACAAAGGTCAAAAAATGATATAAAGTCACCAAA
C